

# 自注意力机制的属性异构信息网络嵌入的商品推荐

王宏琳<sup>1</sup> 杨丹<sup>1</sup> 聂铁铮<sup>2</sup> 寇月<sup>2</sup>

<sup>1</sup>(辽宁科技大学计算机与软件工程学院 辽宁鞍山 114051)

<sup>2</sup>(东北大学计算机科学与工程学院 沈阳 110169)

(aswanghonglin@163.com)

## Attributed Heterogeneous Information Network Embedding with Self-Attention Mechanism for Product Recommendation

Wang Honglin<sup>1</sup>, Yang Dan<sup>1</sup>, Nie Tiezheng<sup>2</sup>, and Kou Yue<sup>2</sup>

<sup>1</sup>(School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, Liaoning 114051)

<sup>2</sup>(School of Computer Science and Engineering, Northeastern University, Shenyang 110169)

**Abstract** Heterogeneous network embedding based recommendation technology has the capability to capture the structural information in the network effectively, thus improving the recommendation performance. However, the existing recommendation technology based on heterogeneous network embedding not only ignores the attribute information of nodes and various types of edge relations between nodes, but also ignores the diverse influences of different nodes' attribute information on recommendation results. To address the above issues, a product recommendation framework based on attributed heterogeneous information network embedding with self-attention mechanism (AHNER) is proposed. The framework utilizes attributed heterogeneous information network embedding to learn the unified low-dimensional embedding representations of users and products. When learning node embedding representation, considering that different attribute information has different effects on recommendation results and different edge relations between nodes reflect users' different preferences for products, self-attention mechanism is exploited to mine the latent information of node attribute information and different edge types and learn attribute embedding representation is learned. Meanwhile, in order to overcome the limitation of traditional dot product method as matching function, the framework also exploits deep neural network to learn more effective matching function to solve the recommendation problem. We conduct extensive experiments on three public datasets to evaluate the performance of AHNER. The experimental results reveal that AHNER is feasible and effective.

**Key words** network embedding; attributed heterogeneous information network; self-attention mechanism; product recommendation; deep neural network

**摘要** 基于异构信息网络嵌入的推荐技术能够有效地捕捉网络中的结构信息,从而提升推荐性能。然而现有的基于异构信息网络嵌入的推荐技术不仅忽略了节点的属性信息与节点间多种类型的边关系,

收稿日期:2021-01-04;修回日期:2021-07-26

基金项目:国家自然科学基金项目(62072084,62072086)

This work was supported by the National Natural Science Foundation of China (62072084, 62072086).

通信作者:杨丹(asyangdan@163.com)

还忽略了节点不同的属性信息对推荐结果不同的影响.为了解决上述问题,提出一个自注意力机制的属性异构信息网络嵌入的商品推荐(attributed heterogeneous information network embedding with self-attention mechanism for product recommendation, AHNER)框架.该框架利用属性异构信息网络嵌入学习用户与商品统一、低维的嵌入表示,并在学习节点嵌入表示时,考虑到不同属性信息对推荐结果的影响不同和不同边关系反映用户对商品不同程度的偏好,引入自注意力机制挖掘节点属性信息与不同边类型所蕴含的潜在信息并学习属性嵌入表示.与此同时,为了克服传统点积方法作为匹配函数的局限性,该框架还利用深度神经网络学习更有效的匹配函数解决推荐问题.AHNER 在 3 个公开数据集上进行大量的实验评估性能,实验结果表明 AHNER 的可行性与有效性.

**关键词** 网络嵌入;属性异构信息网络;自注意力机制;商品推荐;深度神经网络

**中图法分类号** TP391

网络嵌入由于其自身良好的有效性和灵活性,不仅可以解决数据稀疏<sup>[1]</sup>等网络数据存在的问题,在各种下游任务中也已经取得显著的成果,如节点分类<sup>[2]</sup>、链路预测<sup>[3]</sup>、节点聚类<sup>[4]</sup>、网络可视化<sup>[5]</sup>以及推荐<sup>[6]</sup>.

基于网络嵌入的推荐算法根据网络中存在节点类型的数量与边类型的数量分为基于同构信息网络嵌入的推荐算法与基于异构信息网络嵌入的推荐算法 2 种.针对同构信息网络的网络嵌入已经有许多经典的算法,如文献[7-8];针对异构信息网络的网络嵌入也有许多经典的算法,如文献[9-10]等.

然而现实存在的网络节点与边不但分为多种类型,每个节点还拥有自己的属性信息.比如在一个商品推荐网络中,节点类型分为用户和商品 2 种,用户有年龄、性别等属性信息,商品有种类、价格等属性信息,用户与商品之间的关系分为浏览、加入购物车和购买 3 种.用户在选择商品时考虑商品不同的属性信息,这也意味着不同的属性信息对推荐结果有不同程度的影响,用户和商品之间不同类型的边关系也意味着用户对商品不同程度的偏好.但是现有

的网络嵌入方法更关注单一边类型的网络结构,不但忽略了网络中的边分为多种类型,还忽略了节点的属性信息与节点的属性信息对推荐结果的影响.

将拥有多种类型的节点与边且节点拥有属性信息的网络称为属性异构信息网络,如图 1 所示的是一个由用户与商品组成的属性异构信息网络.处理属性异构信息网络的嵌入方法为属性异构信息网络嵌入.在处理属性异构信息网络时会面临一些特别的挑战:

1) 丰富的属性信息.属性异构信息网络中的每个节点都拥有自己丰富的属性信息,而且用户在选择项目时会考虑项目不同的属性信息,如何在考虑不同属性信息对推荐结果影响的同时,学习属性嵌入表示是属性异构信息网络嵌入面临的一大挑战.

2) 网络异构性.现实网络中的节点与节点之间的边分为多种类型,不同类型的边意味着用户对商品不同程度的偏好,如何在学习每种边类型对推荐结果影响的同时,为在每种边关系情况下的节点学习统一的嵌入表示是一大挑战.

3) 融合节点属性嵌入与节点结构嵌入.在网络

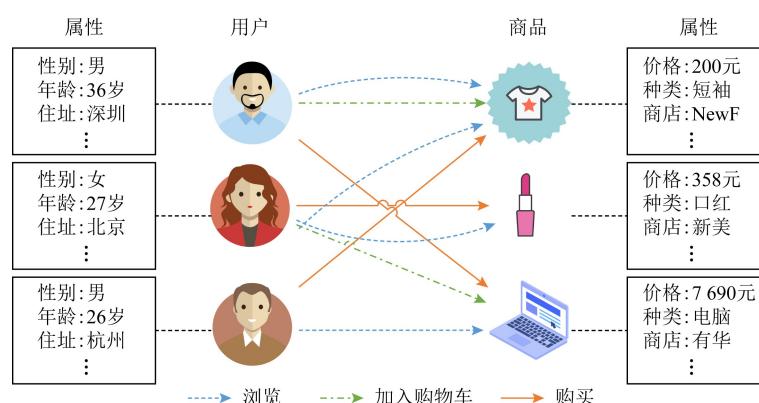


Fig. 1 An example of attributed heterogeneous information network composed of users and products

图 1 用户与商品组成的属性异构信息网络示例图

嵌入中引入属性信息解决推荐问题时,如何融合学习到的节点属性嵌入与节点结构嵌入,学习节点统一低维的嵌入表示是一大挑战。

除这3种属性异构信息网络嵌入所要面临的挑战外,在使用网络嵌入解决推荐问题时,现有的网络嵌入方法大多采用点积方法作为匹配函数计算匹配分数,然后根据匹配分数进行商品排序从而得到商品推荐列表。但是点积方法在解决推荐问题时表现力和求取低秩关系的能力较弱,降低了推荐性能。而基于匹配函数学习的推荐方法相比于学习节点表示更注重如何有效学习匹配函数,这就导致了学习节点表示能力的不足。

为了解决上述问题,本文提出一个自注意力机制的属性异构信息网络嵌入的商品推荐(attributed heterogeneous information network embedding with self-attention mechanism for product recommendation, AHNER)框架。该框架基于自注意力机制利用属性异构信息网络嵌入学习用户嵌入表示与商品嵌入表示,并结合深度神经网络(deep neural network, DNN)学习高效的匹配函数用于解决商品推荐问题。

AHNER在获取节点丰富的属性信息的同时,考虑到不同的属性信息对推荐结果有不同程度的影响,AHNER引入自注意力机制学习每个属性信息的权重,并为每个属性信息学习统一低维的属性嵌入表示。由于属性异构信息网络拥有多种类型的边关系,所以在学习节点的结构嵌入时,AHNER在每种边类型情况下为每个节点学习嵌入表示,然后引入自注意力机制学习每种边类型之间的相互影响,并融合每种边类型情况下的节点嵌入学习最后的节点结构嵌入。最后通过融合机制将学习到的属性嵌入与结构嵌入融合成统一低维的嵌入表示,并将学习到的节点嵌入应用在基于DNN的匹配函数学习模型中,学习匹配分数解决推荐问题。

本文的主要工作概括为4个方面:

1) 强调属性信息在学习节点嵌入时的重要性,引入属性信息学习节点更有效的嵌入表示,并在学习节点属性信息时,引入自注意力机制学习节点属性的权重系数。

2) 详细阐述属性异构信息网络嵌入的学习过程,考虑到不同类型的边关系反映用户对商品不同程度的偏好,引入自注意力机制学习每种边关系的权重,并融合网络结构信息与节点属性信息学习更有效的、统一的、低维的用户嵌入表示与商品嵌入表示。

3) 提出一个推荐框架AHNER用于解决商品推荐问题,并克服传统点积方法在属性异构信息网络嵌入解决推荐问题时的局限性,提高推荐算法的性能。

4) 在3个公共数据集上进行大量的实验,并完成Top-K推荐任务与链路预测任务。实验结果表明,所提出的框架AHNER的性能优于其他方法的性能。

## 1 相关工作

本文的相关工作主要包括基于网络嵌入的推荐技术、基于属性网络嵌入的推荐技术、基于异构信息网络嵌入的推荐技术与基于匹配函数学习的推荐技术这4方面。

### 1.1 基于网络嵌入的推荐技术

现有的基于网络嵌入的推荐技术大致分为基于矩阵分解的网络嵌入推荐技术<sup>[11]</sup>、基于深度学习的网络嵌入推荐技术、基于随机游走的网络嵌入推荐技术3种。早期针对基于矩阵分解的网络嵌入推荐技术研究较多,但是由于对深度神经网络持续的研究发现,深度神经网络可以为网络嵌入寻求一个有效的非线性函数学习模型,如文献[12-14]。与此同时,还发现深度神经网络非常适合学习推荐问题中复杂的匹配函数,因为它们能够逼近任何连续函数<sup>[15]</sup>。例如,文献[16]使用神经网络学习匹配函数;文献[17]采用表示学习和多层感知机(multilayer perceptron, MLP),从用户和项目的输入特征和分类特征中学习匹配函数。基于随机游走的网络嵌入推荐技术的思想是通过邻域的嵌入表示学习推荐网络中节点的嵌入表示,并利用节点的嵌入表示解决推荐问题。例如,word2vec<sup>[18]</sup>是一种通过邻域词的向量构建出中心词的向量的网络嵌入方法。专家发现可以利用word2vec根据用户前后交互的内容推测用户的兴趣,并学习推荐系统中用户、项目的嵌入表示,如文献[19-20]就是使用word2vec的方法解决推荐问题。Deepwalk<sup>[7]</sup>通过游走策略随机选择游走路径,并将随机路径看作一个句子,将节点看作一个词学习节点的嵌入表示。node2vec<sup>[8]</sup>通过使用广度优先(breadth-first search, BFS)和深度优先(depth-first search, DFS)遍历探索不同的邻居节点,从而进一步扩充了Deepwalk。LINE<sup>[21]</sup>是一种能够处理大规模网络的利用邻域相似假设的嵌入方法。基于随机游走的网络嵌入推荐技术通过利用word2vec,

Deepwalk 等基于随机游走的网络嵌入方法解决推荐问题。

## 1.2 基于属性网络嵌入的推荐技术

属性网络<sup>[22]</sup>除了网络的结构信息外,还包括节点的属性信息。属性网络嵌入会将节点的属性信息作为另一种信息源将其映射到一个联合的低维向量空间中学习节点的嵌入表示。MMDW<sup>[23]</sup>是一种利用节点标记信息的半监督网络嵌入方法,它基于深度游走矩阵分解与支持向量机<sup>[24]</sup>结合节点的标签信息学习节点表示。TADW<sup>[25]</sup>是一种基于深度游走矩阵分解结合网络结构与文本信息学习节点嵌入表示的方法,但是它的计算成本高,节点属性也只是简单地合并为无序特征,失去了大量的语义信息。文献[26]将文本内容视为一种特殊的节点,并构建一个增强网络使用逻辑回归(logistic)函数学习节点表示。文献[27]基于光谱技术将标签信息引入到属性网络嵌入中学习节点嵌入表示。文献[28]提出一个可以融合结构信息与属性信息的社交网络通用框架。文献[29]在使用邻居增强自动编码器建模节点属性信息的同时使用跳字模型(skip-gram)汇集邻居节点的特征来学习节点表示。文献[30]利用文本信息学习网络节点表示。

## 1.3 基于异构信息网络嵌入的推荐技术

基于异构信息网络嵌入的推荐技术将用户不同类型的偏好集成到同一公共空间中学习节点统一低维的嵌入表示。例如,metapath2vec<sup>[9]</sup>通过图上的采样路径自动利用邻居结构。文献[31]提出一个推荐系统,它通过获取异构信息网络中的用户隐式偏好进行推荐。文献[32]利用〈user,meta\_path,item〉三元组将基于上下文的元路径合并到交互模型中,并通过MLP对项目进行打分完成Top-K推荐任务。文献[33]通过基于元路径的交互矩阵和注意力机制来识别不同的语义信息并推荐。基于异构信息网络嵌入的推荐技术在推荐效果、多样性等性能都有优秀的表现,但是基于异构信息网络嵌入的推荐技术也存在2点不足:1)忽略属性信息与不同属性信息对推荐结果的影响;2)较少考虑用户与项目之间不同的交互行为对推荐结果的影响。

## 1.4 基于匹配函数学习的推荐技术

基于匹配函数学习的推荐技术更注重学习有效、准确的匹配函数来解决推荐问题。例如,NeuMF<sup>[16]</sup>利用神经网络代替香草矩阵分解中使用的点积来学习匹配函数,还将矩阵分解与多层感知机结合在一个模型中解决推荐问题。文献[34]是NeuMF的一个

变体,它以用户邻居和项目邻居作为输入解决推荐问题。文献[35]使用外部产品操作,以便能够更好地学习推荐系统中用户与项之间的相关性。文献[36]使用基于神经网络的分解机(factorization machines, FM)来避免人工特征工程,解决推荐问题。

基于属性网络嵌入的推荐技术大多局限在同构信息网络中学习节点嵌入表示;基于异构信息网络嵌入的推荐技术不仅忽略了节点丰富的属性信息与不同属性信息对推荐结果的不同影响,还忽略了节点之间不同类型的边对推荐结果的影响。为了解决上述问题,本文使用基于自注意力机制的属性异构信息网络嵌入学习节点的嵌入表示,并为了克服点积方法在推荐问题中存在的局限性,将属性异构信息网络嵌入与DNN结合起来,提出一个通用的推荐框架,在引入节点属性信息的同时学习异构信息网络中节点的嵌入表示,并通过DNN学习高效的匹配函数解决推荐问题。

## 2 预备知识与问题定义

一个网络通常定义为  $G = (V, E)$ , 其中,  $V$  表示节点的集合, 且  $V = \{v_1, v_2, \dots, v_n\}$ ,  $n$  表示节点总数;  $E$  表示边的集合。本文符号具体如表1所示:

Table 1 Symbols and Their Definitions

表1 符号及其含义

符号	含义
$G$	输入的网络
$V, E$	输入网络 $G$ 的节点集合和边集合
$O, R$	输入网络 $G$ 的节点类型集合和边类型集合
$\mathcal{A}$	输入网络 $G$ 的属性集合
$U$	用户集合
$I$	商品集合
$N$	每个节点在每种边类型下邻居节点的集合
$v$	网络中的节点
$r$	一种边类型
$a$	节点的属性信息
$n, m$	网络中节点的数量、边类型的数量
$u, p$	推荐网络中的用户、商品
$d$	整体嵌入维度
$z$	边嵌入的维度
$s, e, \bar{v}$	分别表示节点的结构嵌入、边嵌入、最终节点的嵌入表示

**定义1.** 异构信息网络。一个异构信息网络定义为  $G = (V, E, O, R)$ , 节点与边分别关联一个类型映

射函数  $\Phi: V \rightarrow O$  与  $\Psi: E \rightarrow R$ , 其中  $O$  表示节点类型的集合,  $R$  表示边类型的集合. 如果一个网络中的  $|O| + |R| > 2$ , 则此网络为异构信息网络, 否则为同构信息网络.

**定义 2.** 属性网络. 一个属性网络定义为  $G = (V, E, \mathcal{A})$ . 其中  $\mathcal{A} = \{A_{att}^{(v_1)}, A_{att}^{(v_2)}, \dots, A_{att}^{(v_n)}\}$  是  $G$  的所有节点的属性集合,  $A_{att}^{(v_i)} = \{a_{1,att}^{(v_i)}, a_{2,att}^{(v_i)}, \dots, a_{\tau,att}^{(v_i)}\}$  是节点  $v_i \in V$  的属性集合,  $a_{i,att}^{(v_i)}$  是节点  $v_i$  的一个属性,  $\tau$  是节点  $v_i$  的属性的总数量.

**定义 3.** 属性异构信息网络. 属性异构信息网络定义为  $G = (V, E, \mathcal{A}, O, R)$ . 其中  $E = \bigcup_{r \in R} E_r$ ,  $E_r$  是由边类型为  $r \in R$  的所有边组成边集合. 将每种边类型从整个图中分离出来, 视为子图  $G_r = (V, E_r, \mathcal{A}, O, R)$ . 属性异构信息网络是异构信息网络的一种特例.

**问题 1.** 基于属性异构信息网络嵌入的商品推荐. 给定一个商品推荐的属性异构信息网络  $G_{rec} = (V, E, \mathcal{A}, O, R)$ , 其中节点包含 2 种类型的实体(用户、商品), 边包含用户与商品之间的关系(如购买、加入购物车、浏览等). AHNER 的目标是对任意一个用户  $u \in U$ , 预测用户  $u$  与没交互过的商品  $p \in I$  之间的匹配分数  $rate_{u,p}$ , 并选择 Top-K 个匹配分数最高的商品为用户  $u$  进行推荐.

### 3 推荐框架

本节将详细描述提出的推荐框架 AHNER, 框架结构如图 2 所示. 该框架主要包含 2 部分: 1) 基于商品推荐的属性异构信息网络嵌入; 2) 基于 DNN 的匹配函数学习模型.

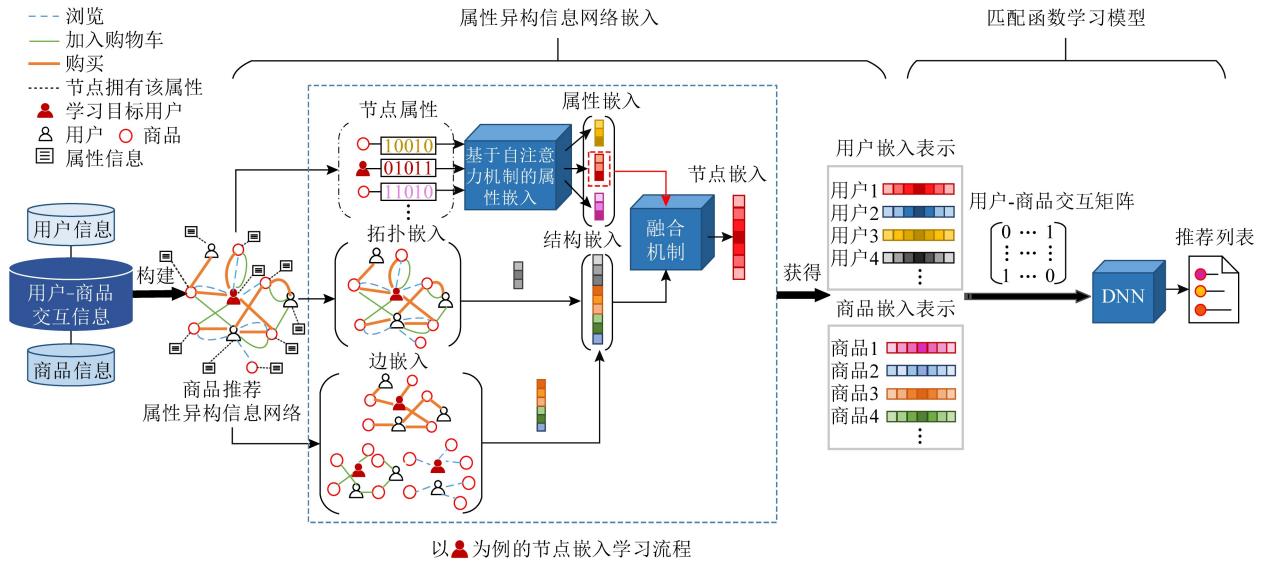


Fig. 2 Overview of AHNER

图 2 AHNER 总体框架图

## 3.1 基于商品推荐的属性异构信息网络嵌入

### 3.1.1 商品推荐属性异构信息网络的构建

构建基于商品推荐的属性异构信息网络时, 节点类型分为用户与商品 2 种. AHNER 使用用户信息作为网络中用户节点的属性信息, 如性别、年龄等, 使用商品信息作为网络中商品节点的属性信息, 如种类、价格等; 又从用户的交互信息中得到节点间的边关系, 其中用户与商品之间的边关系分为 3 种, 分别是浏览、加入购物车与购买. 通过使用 2 组节点集合(即用户和商品)和节点间不同的边关系来连接节点, 从而得到属性异构信息网络.

### 3.1.2 基于自注意力机制的属性嵌入表示学习

在实际推荐问题中, 网络中的节点属性信息是多种多样的, 不同的属性信息对推荐结果产生不同程度的影响. 比如, 在用户选择商品时, 除考虑商品本身的特征外, 还要考虑商品的价格. 在 2 个商品本身特征相似的情况下, 大部分用户会选择优惠力度比较大或价格更便宜的那个商品. 为了建模不同属性信息对推荐结果不同程度的影响, AHNER 引入自注意力机制来学习每个属性的权重系数, 并学习统一的节点属性嵌入. 具体的基于自注意力机制的属性嵌入结构如图 3 所示:

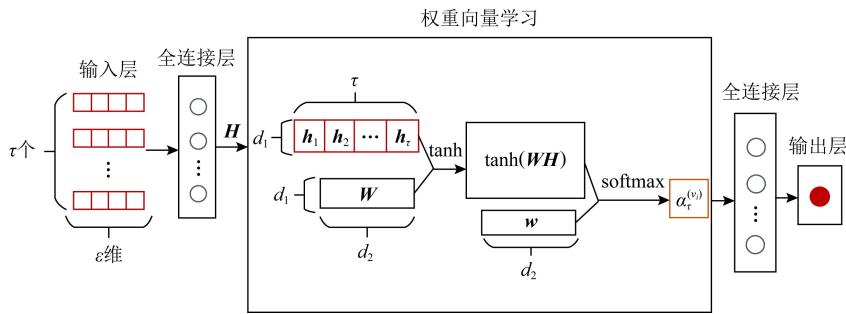


Fig. 3 The architecture of attributed embedding based on self-attention mechanism

图 3 基于自注意力机制的属性嵌入结构图

假设一个节点  $v_i$  共有  $\tau$  个属性信息, 使用 word2vec 方法将节点  $v_i$  的每个属性信息表示成初始的属性向量, 作为基于自注意力机制的属性嵌入的输入数据, 具体表示为

$$\mathbf{A}^{(v_i)} = (\mathbf{a}_1^{(v_i)}, \mathbf{a}_2^{(v_i)}, \dots, \mathbf{a}_{\tau}^{(v_i)}), \quad (1)$$

其中,  $\mathbf{a}_i^{(v_i)}$  是节点  $v_i$  的第  $i$  个属性向量表示,  $\mathbf{A}^{(v_i)}$  是一个由节点  $v_i$  所有属性嵌入向量连接在一起形成的  $\tau \times \epsilon$  的矩阵,  $\tau$  是节点  $v_i$  属性信息的数量,  $\epsilon$  是每个初始属性向量  $\mathbf{a}_i^{(v_i)}$  的维度. 在得到节点  $v_i$  的初始属性向量后, 学习节点  $v_i$  的属性嵌入, 具体学习过程有 3 个步骤.

1) 通过一层全连接层学习输入数据的隐藏状态  $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{\tau})$ , 其中  $\mathbf{h}_t$  表示第  $t$  层隐藏层学习到的隐藏状态.  $\mathbf{H}$  的具体表示为

$$\mathbf{H} = \sigma(\mathbf{W}\mathbf{A}^{(v_i)} + \mathbf{b}), \quad (2)$$

其中,  $\mathbf{H}$  是维度为  $d_1 \times \tau$  的矩阵,  $\mathbf{W}$  与  $\mathbf{b}$  是模型的参数,  $\sigma$  是激活函数.

2) 通过注意力打分函数和 softmax 函数为节点  $v_i$  的每个属性信息计算权重系数. 在 AHNER 中采用加性注意力作为注意力打分函数<sup>[37]</sup>. 节点  $v_i$  的每个属性信息的权重系数表示为

$$\alpha_t^{(v_i)} = \text{softmax}(\mathbf{w}^T \tanh(\mathbf{W}\mathbf{H}))^T, \quad (3)$$

其中,  $\mathbf{W}$  是训练得到的维度为  $d_2 \times d_1$  的权重矩阵,  $\mathbf{w}$  是训练得到的维度为  $d_2$  的权重系数, softmax 确保得到的注意力分数加和为 1. 节点  $v_i$  的全部属性信息的权重系数表示为  $\alpha = \{\alpha_1^{(v_i)}, \alpha_2^{(v_i)}, \dots, \alpha_{\tau}^{(v_i)}\}$ . 得到节点  $v_i$  每个属性信息的属性权重系数后, 将权重系数与  $\mathbf{H}$  加权求和, 得到节点  $v_i$  每个属性信息通过自注意力机制学习得到的属性嵌入  $\mathbf{m}_a$ :

$$\mathbf{m}_a = \alpha_t^{(v_i)} \mathbf{H}. \quad (4)$$

3) 通过全连接层将节点  $v_i$  的每个属性信息的属性嵌入连接到一起, 并输出到输出层得到节点  $v_i$  的所有属性信息的属性嵌入表示.

### 3.1.3 节点结构嵌入表示学习

AHNER 为每种边关系生成一个边嵌入网络  $G_r = (V, E_r, \mathcal{A}, O, R)$ , 然后在每个边嵌入网络中运用基于元路径的随机游走方法<sup>[9]</sup>进行节点采样生成随机序列, 其中元路径的长度设为  $l$ , 元路径表示为

$$\rho: V_1 \rightarrow V_2 \rightarrow \dots \rightarrow V_l \rightarrow \dots \rightarrow V_L.$$

基于元路径的随机游走方法产生的游走流取决于预定义的元路径, 预定义的长度决定游走流的长度. 在游走后, 会得到每一种边类型情况下的基于不同元路径的节点序列. 随机游走的转移概率表示为

$$p(v_j | v_i, \rho) = \begin{cases} \frac{1}{|N_{i,r} \cap V_{t+1}|}, & (v_i, v_j) \in E_r, v_j \in V_{t+1}, \\ 0, & \text{其他,} \end{cases} \quad (5)$$

其中,  $v_i \in V_t$ ,  $N_{i,r}$  表示节点  $v_i$  在边的类型为  $r$  的情况下的邻居节点集合,  $E_r$  表示边类型为  $r$  的边集合.

每个节点在每种边类型的情况下都进行邻居节点采样后, 通过聚合函数聚合邻居节点, 能够得到节点  $v_i$  基于边的类型为  $r$  的情况下的  $k$  阶嵌入表示  $\mathbf{e}_{i,r}^{(k)}$ :

$$\mathbf{e}_{i,r}^{(k)} = \text{aggregator}(\{\mathbf{e}_{j,r}^{(k-1)}, \forall v_j \in N_{i,r}\}), \quad (6)$$

其中,  $\text{aggregator}$  表示聚合函数.

聚合函数有 2 种计算方法<sup>[38]</sup>, 分别是平均池化和最大池化, AHENR 采用平均池化的方法进行计算, 变换后的节点嵌入表示为

$$\mathbf{e}_{i,r}^{(k)} = \sigma(\mathbf{W}_k \cdot \text{mean}(\{\mathbf{e}_{j,r}^{(k-1)}, \forall v_j \in N_{i,r}\})), \quad (7)$$

其中,  $\sigma$  表示激活函数,  $\mathbf{W}_k$  表示权重矩阵.

每个边嵌入的维度为  $z$ , 边类型的总数为  $m$ . 连接不同边类型下的节点  $v_i$  的嵌入表示得到维度  $z \times m$  的矩阵  $\mathbf{M}_i$ :

$$\mathbf{M}_i = (\mathbf{e}_{i,1}, \mathbf{e}_{i,2}, \dots, \mathbf{e}_{i,m}). \quad (8)$$

由于不同类型的边反映用户不同的偏好程度,

所以引入自注意力机制捕捉不同边类型之间的影响因素,学习每种边类型下各个边嵌入的权重系数  $\alpha_{i,r}$ :

$$\alpha_{i,r} = \text{softmax}(\mathbf{w}_r^\top \tanh(\mathbf{W}_r \mathbf{M}_i))^\top, \quad (9)$$

其中,  $\mathbf{w}_r$  和  $\mathbf{W}_r$  是模型训练得到的参数.

最后节点  $v_i$  的结构嵌入表示为

$$\mathbf{s}_{i,r} = \mathbf{g}_i + \delta_r \alpha_{i,r} \mathbf{H}_r^\top \mathbf{M}_i, \quad (10)$$

其中,  $\delta_r$  是边嵌入在整个节点嵌入中重要性的超参数,  $\mathbf{H}_r$  是维度为  $z \times d$  的可训练的转换矩阵,  $\mathbf{g}_i$  是节点  $v_i$  的拓扑嵌入, 拓扑嵌入是提取网络中的基础结构形成的嵌入表示, 它会在每个边嵌入中共享.

### 3.1.4 节点嵌入表示学习

AHNER 采用融合节点结构嵌入与属性嵌入的融合机制学习节点统一的低维嵌入表示, 具体结构如图 4 所示:

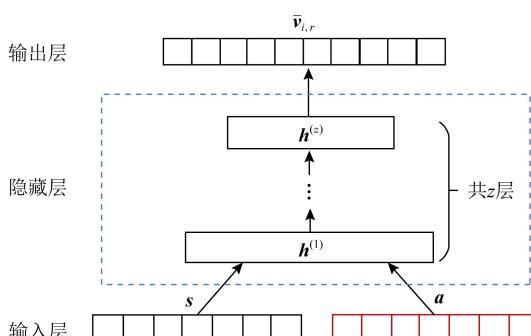


Fig. 4 The architecture of fusion mechanism

图 4 融合机制结构图

输入层的输入数据是节点的属性嵌入表示  $a$  与节点的结构嵌入表示  $s$ , 具体表示为

$$\mathbf{h}^{(0)} = \begin{pmatrix} s \\ a \end{pmatrix}. \quad (11)$$

通过输入层将数据送入隐藏层中, 每个隐藏层表示为  $\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(z)}$ , 具体定义为

$$\mathbf{h}^{(t)} = \theta_t(\mathbf{W}_t \mathbf{h}_{t-1} + \mathbf{b}_t), t=1, 2, \dots, z, \quad (12)$$

其中,  $t$  代表隐藏层的层数,  $\mathbf{W}_t$  与  $\mathbf{b}_t$  分别表示第  $t$  层的权重矩阵与偏置向量,  $\theta_t$  表示第  $t$  层的激活函数.

经过  $z$  个隐藏层, 输出层得到最后的节点表示为

$$\bar{\mathbf{v}}_{i,r} = f_o(a_i) + \delta_r \alpha_{i,r} \mathbf{H}_r^\top \mathbf{M}_i + \beta_r \mathbf{D}_o^\top \mathbf{a}, \quad (13)$$

其中,  $\mathbf{D}_o$  是节点  $v_i$  的特征转换矩阵,  $o$  是节点的类型,  $\beta_r$  是系数. 节点  $v_i$  的拓扑嵌入  $\mathbf{g}_i$  由节点的属性嵌入通过融合机制学到的转化函数  $f_o$  进行计算.

### 3.1.5 属性异构信息网络嵌入优化

假设  $P = (v_1, v_2, \dots, v_l)$  是在边类型为  $r$  的情况下得到的长度为  $l$  的游走序列, 其中  $(v_{t-1}, v_t) \in E_r, t=2, 3, \dots, l$ .

通过基于元路径的随机游走得到的节点  $v_t$  的上下文节点集合定义为

$$C = \{v_k \mid v_k \in P, |k-1| \leq c, t \neq k\}, \quad (14)$$

其中,  $v_k$  是游走序列  $P$  中的节点,  $c$  是上下文窗口的范围.

给定一个节点和元路径, 可以得到在固定大小窗口中的基于共现概率的邻居节点, 则优化目标是最小化式(15):

$$-\log P_\eta(\{v_j \mid v_j \in C\} \mid v_i) = \sum_{v_j \in C} -\log P_\eta(v_j \mid v_i), \quad (15)$$

其中,  $\eta$  表示所有可优化的参数,  $C$  表示上下文节点集合.

遵循 metapath2vec 采用异构 softmax 函数计算  $P_\eta(v_j \mid v_i)$ , 具体表示为

$$P_\eta(v_j \mid v_i) = \frac{\exp(\mathbf{x}_j \cdot \bar{\mathbf{v}}_{i,r})}{\sum_{k \in V_t} \exp(\mathbf{x}_k \cdot \bar{\mathbf{v}}_{i,r})}, \quad (16)$$

其中,  $\mathbf{x}_k$  表示节点  $v_k$  的上下文嵌入表示.

为了高效优化, 引入负采样方法近似  $-\log P_\eta(v_j \mid v_i)$  并得到最终的目标函数表示为

$$X = -\log \sigma(\mathbf{x}_j \cdot \bar{\mathbf{v}}_{i,r}) - \sum_{l=1}^L \mathbb{E}_{v_k^l \sim P_t(v)} [\log \sigma(-\mathbf{x}_k^l \cdot \bar{\mathbf{v}}_{i,r})], \quad (17)$$

其中,  $\sigma(x) = \frac{1}{1+\exp(-x)}$  是 sigmoid 函数,  $P_t(v)$

是在节点  $v_j$  相对应的节点集合  $V_t$  上定义的分布,  $L$  是对应于正训练样本的负样本的数量,  $v_k$  是从  $P_t(v)$  中随机抽取的节点.

### 3.2 基于 DNN 的匹配函数学习模型

为了克服传统点积方法在网络嵌入中求取低秩关系的局限性, AHNER 提出基于 DNN 的匹配函数学习模型, 用于学习高效的匹配函数, 从而提升推荐性能, 具体结构如图 5 所示.

输入层的输入数据是用户与商品形成的交互矩阵和通过属性异构信息网络嵌入学习到的用户嵌入表示与商品嵌入表示. 基于 DNN 的匹配函数学习模型的输入层定义为

$$\mathbf{h}^{(0)} = \begin{pmatrix} \mathbf{p}_u \\ \mathbf{q}_p \end{pmatrix}, \quad (18)$$

其中,  $\mathbf{p}_u$  表示用户  $u$  的嵌入表示,  $\mathbf{q}_p$  表示商品  $p$  的嵌入表示.

通过输入层将数据送入隐藏层中, 经过多个隐藏层学习匹配函数, 每个隐藏层表示为  $\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(z)}$ , 具体定义为

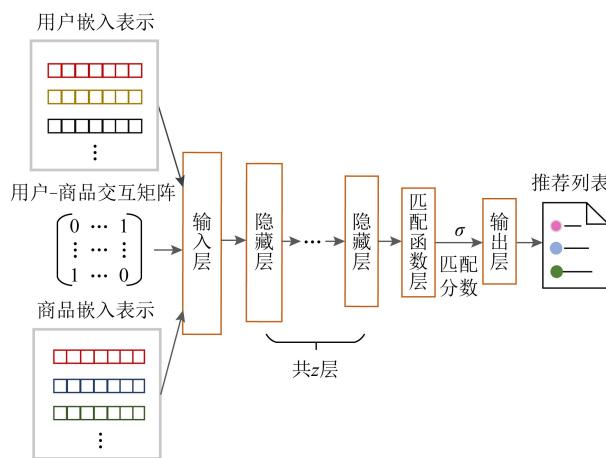


Fig. 5 The architecture of matching function learning model based on DNN

图 5 基于 DNN 的匹配函数学习模型结构图

$$\mathbf{h}^{(t)} = \theta_t(\mathbf{W}^T \mathbf{h}_{t-1} + \mathbf{b}_t), \quad t = 1, 2, \dots, z, \quad (19)$$

其中,  $\mathbf{W}_t$  和  $\mathbf{b}_t$  分别表示第  $t$  层感知器的权重矩阵与偏置向量,  $\theta_t$  表示第  $t$  层的激活函数, 隐藏层的激活函数采用 ReLU 函数.

最后学习到的匹配函数公式为

$$\hat{y}_{ui} = \sigma(\mathbf{W}^T \mathbf{h}_z), \quad (20)$$

其中,  $\sigma$  是激活函数 sigmoid,  $\mathbf{W}$  表示权重矩阵.

## 4 实验评价与分析

本节首先介绍实验数据集、对比方法和评价指标, 然后根据实验数据说明 AHNER 的性能. 本节实验是在 Intel® Xeon® E5-2620 v4@2.10 GHz 的硬件环境下, 基于 python3.6 的 Keras 与 TensorFlow 实现.

### 4.1 实验数据集

AHNER 在 3 个公开数据集 Retailrocket, Amazon Product, YouTube 上完成 Top-K 推荐任务与链路预测任务的性能评价. 表 2 是 3 个实验数据集的统计数据:

Table 2 Statistics of Datasets

表 2 数据集的统计信息

数据集	节点数量	边数量	节点类型	边类型
Retailrocket	18 790	83 429	2	3
Amazon Product	10 099	126 535	1	2
YouTube	2 000	1 320 617	1	5

Retailrocket 数据集是从一个真实的电子商务网站中收集的数据. 其中: 包含“用户”“商品”2 种节

点类型; 节点之间的边类型分为“购买”“加入购物车”“浏览”3 种; 商品节点包括“种类”“价格”等属性信息.

Amazon Product 数据集提供 Amazon 的产品评论和元数据. 由于 Amazon 的产品分为多种类别, 每种类别的产品数量都很大, 所以在实验中只使用电子类产品的元数据. 其中: 节点类型只有“产品”1 种; 节点间的边关系分为“共同浏览”“共同购买”2 种; 产品节点包括“价格”“销售排行”“品牌”“种类”等属性信息.

YouTube 数据集的节点类型只有“用户”1 种; 边的类型分为“联系人”“共享好友”“共享订阅”“共享订阅者”“用户之间共享最喜欢的视频”5 种.

### 4.2 对比方法

实验针对 AHNER 中的属性异构信息网络嵌入和匹配函数学习模型选择不同的对比方法. 针对属性异构信息网络嵌入, 本节选择网络嵌入的方法进行对比实验. 针对匹配函数学习模型, 采用点积方法作为匹配函数进行对比实验.

#### 4.2.1 属性异构信息网络嵌入的对比实验

选择网络嵌入对比方法时, 考虑到网络结构的不同, 分别选择基于同构信息网络的嵌入方法与基于异构信息网络的嵌入方法. 针对同构信息网络, 选择 Deepwalk, node2vec, LINE 作为对比方法. 针对异构信息网络, 选用 metapath2vec 作为对比方法. 除此之外, 选择 AHNER 方法的变形——AHNER-NS 验证 AHNER 基于自注意力机制学习属性嵌入的必要性. 下面逐一介绍各方法.

1) Deepwalk. 在网络图上分离出不同边种类的子图, 为每个子图使用随机游走和 skip-gram 分析节点结构关系并学习节点的向量表示.

2) node2vec. 设计了一种二阶随机游走策略对邻域节点进行采样, 在 BFS 和 DFS 之间平滑插入. 在本节实验中, 参数  $p=2$ , 参数  $q=0.5$ .

3) LINE. 该方法是一种运用一阶相似度与二阶相似度学习节点嵌入表示的同构网络嵌入方法. 在本节实验中, 节点嵌入表示的维度设置为 100.

4) metapath2vec. 该方法能够处理网络中节点的异构性. 但是当网络节点只有一种时, 方法会退化为 Deepwalk. 由于 Amazon Product, YouTube 数据集的节点类型都为 1 种, 所以 metapath2vec 在这 2 个数据集中退化为 Deepwalk. 在 Retailrocket 数据集上, 节点类型包括用户 U 和项目 I, 元路径在实验中设置为 U—I—U 和 I—U—I.

5) AHNER-NS.该方法是 AHNER 的一种变形.在对比实验中使用该方法时,处理离散类型的属性信息使用 one-hot 方法,处理连续型属性信息使用 word2vec 方法.AHNER-NS 与 AHNER 的主要区别是当 AHNER-NS 学习属性嵌入时,不使用自注意力机制.

#### 4.2.2 匹配函数学习的对比实验

针对匹配函数学习的对比,使用 AHNER-ND 作为对比方法.

1) AHNER-ND.该方法是 AHNER 方法的一种变形,在实验中使用传统的点积作为匹配函数代替提出的基于 DNN 的匹配函数学习模型.通过该方法作为对比方法,可以根据实验结果观察到 AHNER 在为匹配函数学习做出改进的重要性.

2) AHNER.该方法是本文提出的商品推荐的方法.在实验中为了公平性,使用与 metapath2vec 相同的元路径方案 U—I—U 和 I—U—I.节点嵌入维度  $d$  设置为 200,边嵌入维度  $z$  设置为 10.每个节点的随机游走次数设置为 20,随机游走的长度设置为 10,窗口大小设置为 5,用于生成节点上下文.每个训练对的负样本数设置为 5,训练 skip-gram 模型的迭代次数设置为 100,epoch 设置为 50,每个边类型下的系数  $\delta_r$  和  $\beta_r$  都设置为 1.AHNER 在 TensorFlow 中使用 Adam 优化器的默认设置,学习速率设置为 0.001.

在实现对比方法时,对比方法中的参数设置与 AHNER 中的参数设置相同.

Table 3 Results of Top-K Recommendation Experiments with Different Methods on Three Datasets

表 3 3 个数据集上不同方法的 Top-K 推荐任务实验结果

方法	Retailrocket		Amazon Product		YouTube	
	HR	NDCG	HR	NDCG	HR	NDCG
Deepwalk	0.2926	0.2004	0.4973	0.3359	0.4381	0.2502
node2vec	0.2937	0.1992	0.4982	0.3345	0.4361	0.2510
LINE	0.2906	0.1883	0.4811	0.3347	0.4383	0.2531
metapath2vec	0.3014	0.2091	0.4978	0.3367	0.4389	0.2539
AHNER-NS	0.3074	0.2108	0.5003	0.3408	0.4681	0.2847
AHNER-ND	0.3114	0.2113	0.5076	0.3411	0.4685	0.2851
AHNER	<b>0.3206</b>	<b>0.2169</b>	<b>0.5162</b>	<b>0.3525</b>	<b>0.4711</b>	<b>0.2880</b>

注:黑体数值表示最优值.

实验结果表明 AHNER 在很大程度上明显优于其他方法.在完成链路预测任务时,各个对比方法采用点积作为匹配函数,具体实验结果如表 4 所示.

#### 4.3 任务与评价指标

实验通过完成 Top-K 推荐任务与链路预测任务来测试 AHNER 的性能.在完成 Top-K 推荐任务时,根据当前和以前会话中用户已购买、加入购物车和浏览的商品来预测用户在当前会话中下一个浏览的商品,并选取 2 个广泛使用的命中率(hit ratio, HR)和归一化折损累计增益(normalized discounted cumulative gain, NDCG)作为 Top-K 推荐任务的评价指标,并以 10 作为推荐列表的截至数量来评估提出推荐框架的推荐性能.HR 评估测试集中的商品是否在所得推荐列表的前 10 个商品中,HR 的值越高说明推荐性能越好;NDCG 评估的是排名质量,NDCG 的值越高说明推荐性能越好.在链路预测任务中,评价指标为常用的受试者工作特征曲线下的面积(area under curve, AUC),AUC 的值越大说明预测效果越好.在本节中,随机选取 3 次实验的实验数据,计算它们的平均数作为本节统计的实验数据.

#### 4.4 实验结果分析

在实验中,按照 90%, 7%, 3% 的比例随机选择训练集、测试集与验证集.在完成 Top-K 推荐任务时,为了更方便地观察 AHNER 的性能,在实现 Deepwalk, node2vec, LINE, metapath2vec, AHNER-NS 时,将这 5 种方法学到的嵌入表示输入基于 DNN 的匹配函数学习模型中求出相应评价指标的值,在实现 AHNER-ND 时采用传统的点积方法作为匹配函数得到推荐结果并计算相应的评价指标的值,具体实验结果如表 3 所示:

从表 4 发现:

1) 异构信息网络嵌入引入属性信息会显著提高推荐性能.通过表 3 与表 4 中的实验结果可以发现,

引入属性信息会显著提高网络嵌入方法的性能,将更多的属性信息纳入推荐系统将提高总体的推荐性能;与其他只关注用户与项目的方法相比,AHNER通过结合属性信息更能捕获节点间的相互关系,从而提升推荐性能。从表4中可以观察到在链路预测问题上AHNER的性能也是优于其他对比方法的,这也体现出AHNER的适用性与普遍性。

**Table 4 AUC of Link Prediction Experiments with Different Methods on Three Datasets**

**表 4 3个数据集上不同方法的链路预测实验的 AUC**

方法	Retailrocket	Amazon Product	YouTube
Deepwalk	59.53	94.36	70.89
node2vec	63.36	94.51	70.96
LINE	53.90	82.08	70.83
metapath2vec	61.18	94.04	71.19
AHNER-NS	68.82	94.28	82.93
<b>AHNER</b>	<b>71.08</b>	<b>96.53</b>	<b>84.51</b>

注:黑体数值表示最优值。

2) 使用基于自注意力机制的属性嵌入表示学习能够提高推荐性能,通过对比AHNER-NS与AHNER的实验结果可以发现,AHNER通过自注意力机制学习带有权重的属性嵌入后,能捕获更多属性信息和节点之间的相互作用关系,提升了推荐系统的性能。数据集提供的属性信息越多,提升性能的幅度越大,Retailrocket在这3个数据集中提供的属性信息最多,所以提升的性能幅度更大,在HR@10上提升4.3%,在NDCG@10上提升2.9%。

3) 基于DNN的匹配函数学习模型比基于点积的方法更有优势。通过对比AHNER-ND与AHNER的实验结果可以发现,使用深度神经网络作为匹配函数的方法比使用点积作为匹配函数的方法取得了更好的结果。这种性能增益是合理的,因为神经单元增加了模型的容量。结果表明,通过使用参数神经网络替换传统的点积方法作为匹配函数来解决推荐问题,能够更好地在推荐系统中进行学习和交互建模,提升推荐性能。

#### 4.4.1 冷启动场景的实验结果

为了研究AHNER在冷启动场景中的性能,本节改变Retailrocket数据集中训练集的比例,将用于实验的训练集比例从90%逐渐降低至20%,然后利用AHNER解决链路预测问题,具体实验结果如表5所示:

**Table 5 AUC of Cold Start on Retailrocket Dataset**

**表 5 Retailrocket 数据集上冷启动场景的 AUC**

方法	训练集比例/%			
	90	70	50	20
Deepwalk	59.53	59.11	58.11	56.91
metapath2vec	61.18	60.65	58.97	57.07
AHNER-NS	68.82	68.56	68.24	67.61
<b>AHNER</b>	<b>71.08</b>	<b>70.70</b>	<b>70.35</b>	<b>69.88</b>

注:黑体数值表示最优值。

通过分析实验结果,在训练集比例从90%逐渐降低至20%时,2个对比方法的AUC值分别下降4.6%与7.2%,但AHNER的性能仅下降1.7%~1.8%。这表明AHNER即使用户与商品之间的交互稀疏,也能保持预测性能。

#### 4.4.2 属性权重可视化

本节在Amazon数据集中随机选择5个商品进行属性权重系数可视化,结果如图6所示:

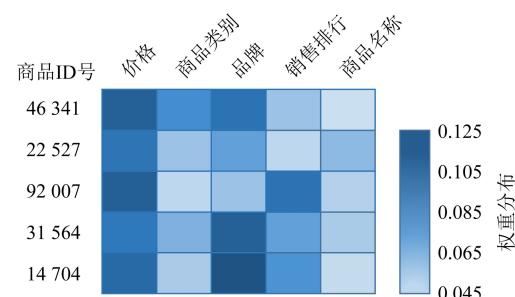


Fig. 6 The attribute weight coefficient of different products

图 6 不同商品的属性权重系数图

图6中所示的权重系数都是由基于自注意力机制的属性嵌入表示学习出的商品属性的权重系数,其中每一行记录一个商品(由商品ID号表示)的权重系数。从实验结果中可以观察出:

1) 商品不同的属性信息,权重系数也是不同的。这一现象与本文的猜想是一致的,即不同的属性信息对最终嵌入表示的贡献是不同的。

2) “价格”属性的权重系数在每个商品的属性信息中占比较大。这一发现与用户在实际推荐系统中的消费行为相一致。除此之外,“品牌”属性的权重系数也比较大,这种现象也反映出用户的消费行为规律,即用户倾向于购买同一品牌的商品,以方便降低价格或领取优惠。

3) 对于不同的商品,用户倾向考虑的属性信息也会不同。比如商品ID为14 704的商品,用户除考

虑商品本身的特点外,相比于“销售排行”会更加考虑商品的品牌,所以“品牌”的权重占比会比较大.而在商品 ID 为 92 007 的商品中,用户会相比于“品牌”属性更多地考虑商品的“销售排行”属性.

#### 4.4.3 参数分析

本节将讨论 AHNER 的参数问题,主要涉及在进行属性异构信息网络嵌入学习时所采用的边嵌入维度  $z$ .通过保持其他设置不变,改变边嵌入维度  $z$  的值,观察 AHNER 的性能变化,图 7 给出了调整边嵌入维度  $z$  时,AHNER 在 Retailrocket 数据集上性能的表现.

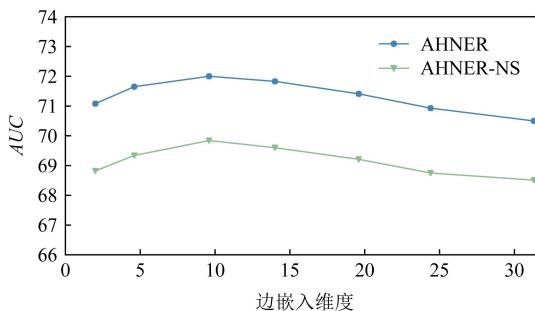


Fig. 7 Parameter performance changes of AHNER on Retailrocket dataset

图 7 Retailrocket 数据集上 AHNER 参数性能变化图

从图 7 中可以观察到,边嵌入维度对 AHNER 的推荐性能影响较弱,当维度取值在 10 左右时 AHNER 的性能最好,当维度取值大于 10 时性能平缓下降,当维度取值小于 10 时性能平稳上升.

## 5 总 结

本文提出一个通用的基于自注意力机制的属性异构信息网络嵌入的商品推荐框架——AHNER,该框架通过基于自注意力机制的属性异构信息网络嵌入学习用户与商品的嵌入表示并解决推荐问题.AHNER 在学习嵌入表示时充分利用节点的属性信息,学习不同属性信息的权重系数,并学习节点的属性嵌入表示.在学习节点的结构嵌入时,使用自注意力机制解决属性异构信息网络中多种边类型的问题,并为其学习统一的节点嵌入表示.然后融合节点的属性嵌入与节点的结构嵌入为节点学习整体统一的、低维的节点嵌入表示,并在推荐过程中克服点积作为匹配函数的局限性,利用 DNN 学习高效的匹配函数并解决推荐问题.AHNER 在 3 个公开数据

集上进行实验,实验结果证明 AHNER 比以前的方法取得了更好的性能.在未来的工作中:1)研究如何引入多模态的商品描述信息,如用户对商品的评论、商品图片、视频等来丰富网络中的节点属性信息,从而提高推荐性能;2)研究如何引入知识图谱来增强推荐系统的可解释性.

**作者贡献声明:** 王宏琳负责方法的设计与实现,以及论文的撰写; 杨丹负责确定研究思路和全文结构设计,并指导论文撰写; 聂铁铮指导实验方案的实现; 寇月负责审阅和完善论文内容.

## 参 考 文 献

- [1] Chen Xuening, Liu Hanwen, Yang Dan. Improved LSH for privacy-aware and robust recommender system with sparse data in edge environment [J/OL]. EURASIP Journal on Wireless Communications and Networking, 2019. [2020-02-01]. <https://doi.org/10.1186/s13638-019-1478-1>
- [2] Ye Zhonglin, Zhao Haixing, Zhang Ke, et al. Network representation learning using the optimizations of neighboring vertices and relation model [J]. Journal of Computer Research and Development, 2019, 56(12): 2562–2577 (in Chinese)  
(冶忠林, 赵海兴, 张科, 等. 基于邻节点和关系模型优化的网络表示学习[J]. 计算机研究与发展, 2019, 56(12): 2562–2577)
- [3] Wu Yao, Shen Derong, Kou Yue, et al. Heterogeneous information networks embedding based on multiple meta-graph fusion [J]. Journal of Computer Research and Development, 2020, 57(9): 1928–1938 (in Chinese)  
(吴瑶, 申德荣, 寇月, 等. 多元图融合的异构信息网嵌入[J]. 计算机研究与发展, 2020, 57(9): 1928–1938)
- [4] Chang Shiyu, Han Wei, Tang Jiliang, et al. Heterogeneous network embedding via deep architectures [C] //Proc of the 21st ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2015: 119–128
- [5] Pan Shirui, Wu Jia, Zhu Xingquan, et al. Tri-party deep network representation [C] //Proc of the 25th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2016: 1895–1901
- [6] Zhao Yu, Liu Zhiyuan, Sun Maosong. Representation learning for measuring entity relatedness with rich information [C] //Proc of the 24th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2015: 1412–1418
- [7] Perozzi B, Ai-rfou R, Skiena S. Deepwalk: Online learning of social representations [C] //Proc of the 20th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701–710

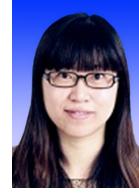
- [8] Grover A, Leskovec J. node2vec: Scalable feature learning for networks [C] //Proc of the 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 855–864
- [9] Dong Yuxiao, Chawla N V, Swami A. metapath2vec: Scalable representation learning for heterogeneous networks [C] //Proc of the 23rd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2017: 135–144
- [10] Shi Chuan, Hu Binbin, Zhao Xin Wayne, et al. Heterogeneous information network embedding for recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(2): 357–370
- [11] Ou Mingdong, Cui Peng, Pei Jian, et al. Asymmetric transitivity preserving graph embedding [C] //Proc of the 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1105–1114
- [12] Wang Daixin, Cui Peng, Zhu Wenwu. Structural deep network embedding [C] //Proc of the 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1225–1234
- [13] Cao Shaosheng, Lu Wei, Xu Qiongkai. Deep neural networks for learning graph representations [C] //Proc of the 30th AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2016: 1145–1152
- [14] Wang Suhang, Tang Jiliang, Aggarwal C, et al. Signed network embedding in social media [C] //Proc of the 17th SIAM Int Conf on Data Mining. Philadelphia, PA: SIAM, 2017: 327–335
- [15] Hornik K, Stinchcomb M, White H. Multilayer feedforward networks are universal approximators [J]. Neural Networks, 1989, 2(5): 359–366
- [16] He Xiangnan, Liao Lizi, Zhang Hanwang, et al. Neural collaborative filtering [C] //Proc of the 26th Int Conf on World Wide Web. New York: ACM, 2017: 173–182
- [17] Cheng Heng-Tze, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems [C] //Proc of the 1st Workshop on Deep Learning for Recommender Systems. New York: ACM, 2016: 7–10
- [18] Mikolov T, Sutskever I, Chen Kai, et al. Distributed representations of words and phrases and their compositionality [C] //Proc of the 26th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2013: 3111–3119
- [19] Grbovic M, Radosavljevic V, Djuric N, et al. E-commerce in your inbox: Product recommendations at scale [C] //Proc of the 21st ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2015: 1809–1818
- [20] Grbovic M, Cheng Haibin. Real-time personalization using embeddings for search ranking at Airbnb [C] //Proc of the 24th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2018: 311–320
- [21] Tang Jian, Qu Meng, Wang Mingzhe, et al. LINE: Large-scale information network embedding [C] //Proc of the 24th Int Conf on World Wide Web. Berlin: Springer, 2015: 1067–1077
- [22] Chen Yiqi, Qian Tieyun, Li Wanli, et al. Exploiting composite relation graph convolution for attributed network embedding [J]. Journal of Computer Research and Development, 2020, 57(8): 1674–1682 (in Chinese)  
(陈亦琦, 钱铁云, 李万理, 等. 基于复合关系图卷积的属性网络嵌入方法[J]. 计算机研究与发展, 2020, 57(8): 1674–1682)
- [23] Tu Cunchao, Zhang Weicheng, Liu Zhiyuan, et al. Maxmargin Deepwalk: Discriminative learning of network representation [C] //Proc of the 25th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2016: 3889–3895
- [24] Hearst M A, Dumais S T, Osman E, et al. Support vector machines [J]. IEEE Intelligent Systems and Their Applications, 1998, 13(4): 18–28
- [25] Yang Cheng, Liu Zhiyuan, Zhao Deli, et al. Network representation learning with rich text information [C] //Proc of the 24th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2015: 2111–2117
- [26] Sun Xiaofei, Guo Jiang, Ding Xiao, et al. A general framework for content-enhanced network representation learning [EB/OL]. (2016-07-24) [2020-10-13]. <https://arxiv.org/abs/1610.02906>
- [27] Huang Xiao, Li Jundong, Hu Xia. Label informed attributed network embedding [C] //Proc of the 10th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2017: 731–739
- [28] Liao Lizi, He Xiangnan, Zhang Hanwang, et al. Attributed social network embedding [J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 30(12): 2257–2270
- [29] Zhang Zhen, Yang Hongxia, Bu Jiajun, et al. ANRL: Attributed network representation learning via deep neural networks [C] //Proc of the 27th Int Joint Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2018: 3155–3161
- [30] Tu Cunchao, Liu Han, Liu Zhiyuan, et al. CANE: Context-aware network embedding for relation modeling [C] //Proc of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2017: 1722–1731
- [31] Yu Xiao, Ren Xiang, Sun Yizhou, et al. Personalized entity recommendation: A heterogeneous information network approach [C] //Proc of the 7th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2014: 283–292
- [32] Hu Binbin, Shi Chuan, Zhao Xin Wayne, et al. Leveraging meta-path based context for top- $n$  recommendation with a neural co-attention model [C] //Proc of the 24th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2018: 1531–1540

- [33] Wang Zekai, Liu Hongzhi, Du Yingpeng, et al. Unified embedding model over heterogeneous information network for personalized recommendation [C] //Proc of the 28th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2019: 3813–3819
- [34] Bai Ting, Wen Jirong, Zhang Jun, et al. A neural collaborative filtering model with interaction-based neighborhood [C] // Proc of the 2017 ACM on Conf on Information and Knowledge Management. New York: ACM, 2017: 1979–1982
- [35] He Xiangnan, Du Xiaoyu, Wang Xiang, et al. Outer product-based neural collaborative filtering [C] //Proc of the 27th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2018: 2227–2233
- [36] Guo Huirong, Tang Ruiming, Ye Yunming, et al. DeepFM: A factorization-machine based neural network for CTR prediction [C] //Proc of the 26th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2017: 1725–1731
- [37] Lin Zhouhan, Feng Minwei, Santos C, et al. A structured self-attentive sentence embedding [C/OL] //Proc of the 5th Int Conf on Learning Representations. 2017 [2020-03-20]. <http://export.arxiv.org/abs/1703.03130>
- [38] Hamilton L W, Ying Zhitao, Leskovec J. Inductive representation learning on large graphs [C] //Proc of the 31st Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2017: 1025–1035



**Wang Honglin**, born in 1996. Master. Student member of CCF. Her main research interests include recommendation system and data integration.

王宏琳,1996 年生.硕士,CCF 学生会员.主要研究方向为推荐系统和数据集成.



**Yang Dan**, born in 1978. PhD, professor. Member of CCF. Her main research interests include data integration and recommendation system.

杨丹,1978 年生.博士,教授,CCF 会员.主要研究方向为数据集成和推荐系统.



**Nie Tiezheng**, born in 1980. PhD, associate professor. Senior member of CCF. His main research interests include data quality and data integration.

聂铁铮,1980 年生.博士,副教授,CCF 高级会员.主要研究方向为数据质量和数据集成.



**Kou Yue**, born in 1980. PhD, associate professor. Senior member of CCF. Her main research interests include recommendation system and Web mining.

寇月,1980 年生.博士,副教授,CCF 高级会员.主要研究方向为推荐系统和网络挖掘.